Submitted to the Statistical Science

Measurement error in GLMMs with INLA

Stefanie Muff*, Andrea Riebler*, Håvard Rue, Philippe Saner, and Leonhard Held

Abstract. Measurement error (ME) theory has largely been neglected in many applied sciences, although central variables are often difficult to measure and may thus contain considerable error. To account for ME in explanatory variables, Bayesian approaches provide a flexible framework. However, given the analytic intractability of the posterior distribution, model inference so far has to be performed via time-consuming and complex Markov chain Monte Carlo implementations. In this paper we extend the Integrated nested Laplace approximations (INLA) approach to formulate Gaussian ME models in generalized linear mixed models. We present three applications, and illustrate how parameter estimates are obtained for common ME models, such as the classical and Berkson error model including heteroscedastic variances. We conclude that the solution presented here provides an attractive alternative to likelihood-based approaches and hope it will stimulate the greater use of Bayesian methods for ME modelling. To illustrate the practical feasibility, R-code is provided.

Key words and phrases: Bayesian analysis, Berkson error, Classical error, Integrated nested Laplace approximation, Measurement error.

1. INTRODUCTION

The existence and the effects of measurement error (ME) in statistical analyses have been recognized and discussed for more than a century, see for example Pearson (1902); Wald (1940); Berkson (1950); Fuller (1987); Carroll et al. (2006). The sources of ME are manifold and imply much more than just instrumental imprecision in the measurement of physical variables, such as length, weight etc., but may include for instance biases due to preferential sampling, incomplete observations or misclassification.

Unless the ME is negligible, modeling it is crucial. If ME is ignored, parameter estimates and confidence intervals in statistical models often suffer from serious

Division of Biostatistics, Institute of Social and Preventive Medicine, University of Zurich, Switzerland (e-mail: stefanie.muff@uzh.ch; andrea.riebler@uzh.ch; leonhard.held@ifspm.uzh.ch).
Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Switzerland (e-mail: stefanie.muff@uzh.ch; philippe.saner@uzh.ch).
Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway (e-mail: havard.rue@math.ntnu.no).
*These two authors contributed equally to this work.

biases. If a regression model is multivariate and some covariates can be measured with and some without error, even the effects of the error-free measured covariates can be biased, where the direction of the bias depends on the correlation among covariates (Carroll, Gallo and Gleser, 1985; Gleser, Carroll and Gallo, 1987). Moreover, ignoring ME may cause a loss of power for detecting signals and connections among variables, and may mask important features of the data. Given these facts, it is surprising that ME is often completely ignored or not treated properly. One reason might be that the statistical standard literature often pays very little attention to this aspect, although the problems have been recognized for a long time.

For successful error-correction both the amount of error (i.e., the error variance) and the error model need to be specified correctly. Hence, information about the underlying measurement process is essential. Possible errors must be identified early in a study and the entire data-collection process should be driven by such considerations. In the last decades, several approaches to model and correct for ME have been proposed, such as method-of-moments corrections (Fuller, 1987), simulation extrapolation (SIMEX) (Cook and Stefanski, 1994), regression calibration (Carroll and Stefanski, 1990; Gleser, 1990), or Bayesian analyses (Clayton, 1992; Stephens and Dellaportas, 1992; Richardson and Gilks, 1993; Dellaportas and Stephens, 1995; Gustafson, 2004). A thorough overview of current state-of-the-art methods is given in the books of Carroll et al. (2006) and Buonaccorsi (2010).

In this paper, we focus on Bayesian approaches, as they provide a flexible framework and are thus well suited to treat ME. One of the main advantages of is that prior knowledge, and in particular prior uncertainty, e.g., in variance estimates, can be incorporated in the model. While frequentist approaches require to fix one or several parameters to guarantee identifiability, the Bayesian setting allows to represent uncertainty with suitable prior distributions. Note that the fixation of parameters in a frequentist setting corresponds to a Bayesian point prior.

Up to now, posterior marginal distributions of such errors-in-variables models have been estimated by employing a Markov chain Monte Carlo (MCMC) sampler, see for example Stephens and Dellaportas (1992) or Richardson and Gilks (1993). However, case-specific implementation may be challenging, MCMC is time-consuming and its analysis and interpretation requires diagnostic tools. Generic software like WinBugs (Lunn et al., 2000), OpenBugs (Lunn et al., 2009), or MCMC samplers in `R`, such as MCMCpack (Martin, Quinn and Park, 2011) or LaplacesDemon (Statisticat and LLC., 2013), might be used, but they suffer from the same drawbacks as any MCMC techniques.

Recently, an alternative to MCMC has been proposed to estimate posterior marginals by integrated nested Laplace approximations (INLA) for the class of latent Gaussian models (Rue, Martino and Chopin, 2009). INLA provides accurate approximations avoiding time-consuming sampling. Due to its flexibility in the choice of likelihood functions and latent models, INLA is an appealing alternative to likelihood-based inference in particular for generalized linear mixed models (GLMMs) (Fong, Rue and Wakefield, 2010). The INLA approach is implemented in `C` and easy to use under Linux, Windows and Macintosh via a freely available `R`-interface (R Core Team, 2012). The `R`-package `r-inla` can be down-

loaded from www.r-inla.org. It allows to easily specify models in a modular way, where different types of regression models can be combined with different types of error models. Moreover, it is straightforward to incorporate random effects, such as independent or conditional autoregressive (CAR) models to account for spatial structure, which is of importance in several settings (Bernardinelli et al., 1997). Here, we used the `r-inla` version updated on Feb 10, 2013.

In this paper we extend the INLA framework to the most common Gaussian ME models, namely the classical and the Berkson ME models, which are suitable for continuous error-prone covariates. To facilitate the usage of the INLA-package with the new features, R-code is provided in the Supplementary Material. We hope that the solution presented here will increase the use of ME thinking in practice and stimulates the greater use of Bayesian methods in ME modeling.

Section 2 introduces three applications from the biological/medical field containing: a linear regression combined with heteroscedastic classical error, a logistic model with an exposure model suffering from classical error, and an overdispersed Poisson regression model with Berkson error. In Section 3 we will review the classical and Berkson ME models and their effects. Bayesian analysis with INLA is introduced in Section 4, where we will describe how to use this framework for model inference in the presence of classical and Berkson ME. Section 5 presents modeling details and the results of the three applications analyzed with both INLA and MCMC. Finally, we provide a discussion and outlook in Section 6.

2. EXAMPLES OF MEASUREMENT ERROR PROBLEMS

2.1 Inbreeding in Swiss ibex populations

We analyzed data described by Bozzuto et al. (2013) on 26 Alpine ibex populations in Switzerland, some of them monitored over the last 100 years. The study aimed to quantify the effect of inbreeding on populations' intrinsic growth rates $\boldsymbol{y}$. The intrinsic growth rate is the theoretical maximal rate of increase of a population, if there are no density-dependent effects. The inbreeding coefficient x_i of population i (often denoted as f) is a quantity between 0 and 1, with larger values indicating that the population is stronger inbred. This quantity has been estimated from genotype analyses at 37 neutral microsatellite loci. To this end, a Bayesian analysis was employed to derive the estimates w_i and error variances for each population i. If a linear regression model $\boldsymbol{y} = \beta_0 + \beta_x \boldsymbol{x} + \boldsymbol{\varepsilon}$ is fitted with $\boldsymbol{w}$ instead of $\boldsymbol{x}$, the absolute value of the slope parameter $|\beta_x|$ is underestimated ($\hat{\beta}_x = -2.59$, 95% CI: $[-5.44, 0.25]$). Accounting for ME in $\boldsymbol{w}$, the effect of inbreeding on population growth dynamics is more pronounced ($\hat{\beta}_x = -3.03$, 95% CI: $[-6.18, -0.08]$).

2.2 Influence of systolic blood pressure on coronary heart disease

The Framingham heart study is a large cohort study carried out between 1973 and 1986, and aimed to understand the factors leading to coronary heart disease (Kannel, Neaton and Wentworth, 1986). The data used here were originally presented in MacMahon et al. (1990). The main predictor in this logistic regression model is systolic blood pressure (SBP). As in Carroll et al. (2006, Section 9.10), we analyzed data of $n = 641$ males considering smoking status $\boldsymbol{z}$ and $\boldsymbol{x} = \log(\text{SBP} - 50)$ as predictors. The transformation of SBP was originally proposed in Cornfield (1962), and used by Carroll et al. (1984), Carroll et al. (1996),

and in Carroll et al. (2006). Since it is impossible to measure the long-term SBP, measurements at single clinical visits had to be used as a proxy. Note that, due to daily variations or deviations in the measurement instrument, the single-visit measures might considerably differ from the long-term blood pressure (Carroll et al., 2006). Hence, the ME in SBP has been a concern for many years in this study. Importantly, the magnitude of the error could be estimated, as SBP had been measured twice at different examinations ($\boldsymbol{w}_1$, $\boldsymbol{w}_2$). A naive approach ignoring ME would fit a logistic regression against the indicator of coronary heart disease $\boldsymbol{y} \in \{0,1\}^n$

$$\text{logit}\left[P(\boldsymbol{y}=1 \mid \boldsymbol{x}, \boldsymbol{z})\right] = \beta_0 + \beta_x \boldsymbol{x} + \beta_z \boldsymbol{z} \ ,$$

where the true covariate $\boldsymbol{x}$ is replaced by the average of the two measurements $\overline{\boldsymbol{w}} = (\boldsymbol{w}_1 + \boldsymbol{w}_2)/2$. The slope β_x is attenuated in this naive regression ($\hat{\beta}_x = 1.66$, 95% CI: $[0.69, 2.63]$) compared to the estimate obtained with error modeling ($\hat{\beta}_x = 1.93$, 95% CI: $[0.76, 3.05]$).

2.3 Seedling growth across different light conditions

In a shadehouse experiment, described in Paine et al. (2012), the impact of shading (dark, middle, light) and defoliation (0, 25%, 50%, 75% reduction of leaf surface) on plant growth in the Malaysian rainforest was investigated. The number of new leaves per plant after a four months growth phase was counted and used as the response variable for plant growth. Here, we analyzed 60 seedlings from the species *Shorea fallax*, from which 20 individuals were grown each under dark, middle, and light shading conditions. There were five shadehouses for each of the three shading conditions, and each shadehouse contained four individuals. Each seedling in a shadehouse was exposed to a different degree of defoliation treatment, compare Figure 1. In experimental studies in ecology, it is common practice that the value for the target light intensity is assigned to all replicates within a treatment class (i.e., dark, middle, light). However, it is obvious that due to external conditions the actual observed light availability might considerably vary from the target value within replicates.

In this application we can quantify the introduced error, as the actual experimentally created light availability was measured for each replicate. Let the covariate $\boldsymbol{x} = \log(\%\text{light})$ denote the correct light intensities, and $\boldsymbol{w}$ the respective target values. The selected regression model is Poisson and includes an unstructured random term to account for potential overdispersion. In contrast to the above examples 2.1 and 2.2, where the bias induced by including $\boldsymbol{w}$ instead of $\boldsymbol{x}$ in the regression attenuates the parameter estimates, the respective regression coefficient is overestimated here. We will illustrate in Section 5.3 how this bias becomes larger when the error variance is (artificially) increased.

3. MODELS FOR MEASUREMENT ERROR IN REGRESSION

3.1 The generalized linear model

Assume we have n observations in a generalized linear model (GLM). The data are given as $(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{x})$, with $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ denoting the response, $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)^\top$ a covariate matrix of dimension $n \times p$ for p error-free covariates, and $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ a single error-prone covariate whose true values are unobservable. The generalization to multiple error-prone covariates is straightforward.

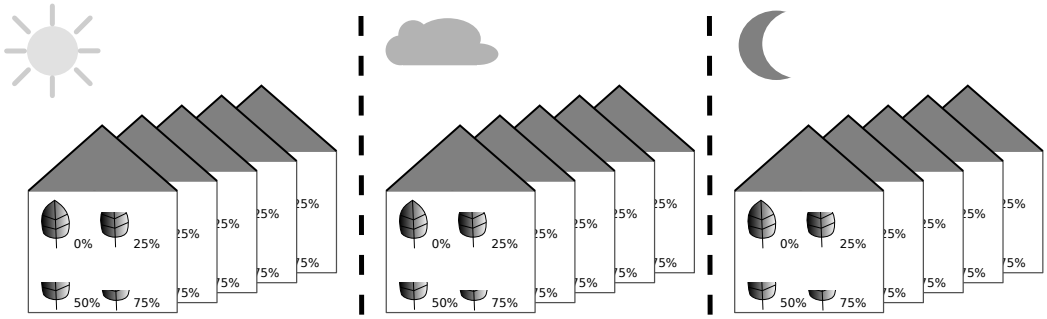

FIG 1. *Illustration of the shadehouse experiment. There were five shadehouses per light condition and each shadehouse contained four seedlings. All seedlings in a shadehouse were exposed to a different defoliation treatment, 0% indicating that the leaves were not cut, 25% that one fourth of each leaf was cut, etc.*

Suppose $\boldsymbol{y}$ is of exponential family form with mean $\mu_i = E(y_i \,|\, \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\beta})$, linked to the linear predictor η_i via

$$\mu_i = h(\eta_i) \tag{1a}$$

$$\eta_i = \beta_0 + \beta_x x_i + \boldsymbol{\beta}_z^\top \boldsymbol{z}_i \ . \tag{1b}$$

Here, $h(\cdot)$ is a known monotonic inverse link (or response) function, β_0 denotes the intercept, β_x the fixed effect for the error-prone covariate $\boldsymbol{x}$, and $\boldsymbol{z}_i$ is $p \times 1$ with a corresponding vector $\boldsymbol{\beta}_z$ of fixed effects. This GLM is extended to a generalized linear mixed model (GLMM) by adding normally distributed random effects on the linear predictor scale (1b).

Let $\boldsymbol{w} = (w_1, \ldots, w_n)^\top$ denote the observed version of the true, but unobservable covariate $\boldsymbol{x}$. We distinguish between two different ME processes: the classical error model and the Berkson error model (Berkson, 1950). The graphical structure of these models is very similar, compare Figure 2, but the effects caused by these models are fundamentally different.

3.2 Classical measurement error model

In the classical error model it is assumed that the covariate $\boldsymbol{x}$ can be observed only via a proxy $\boldsymbol{w}$, such that, in vector notation,

$$\boldsymbol{w} = \boldsymbol{x} + \boldsymbol{u} \ ,$$

with $\boldsymbol{u} = (u_1, \ldots, u_n)^\top$. Classical ME is unbiased and throughout the paper the components of the error vector $\boldsymbol{u}$ are assumed to be independent and normally distributed with mean zero and variance τ_u^{-1}, i.e. $\text{Cov}(u_i, u_j) = 0$ for $i \neq j$. Note that in the following we parameterize the normal distribution with mean and precision (or precision matrix in the multivariate context), rather than using the variance or covariance matrix. Thus we write $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \tau_u \boldsymbol{I})$.

We assume that the error term $\boldsymbol{u}$ is independent of the true covariate $\boldsymbol{x}$, but also independent of any other covariates $\boldsymbol{z}$ and the response $\boldsymbol{y}$. This implies a non-differential ME, meaning that $\boldsymbol{y}$ and $\boldsymbol{w}$ are conditionally independent given $\boldsymbol{z}$ and $\boldsymbol{x}$. In most applications this assumption is plausible as it implies that, given the true covariate $\boldsymbol{x}$ and covariates $\boldsymbol{z}$, no additional information about the

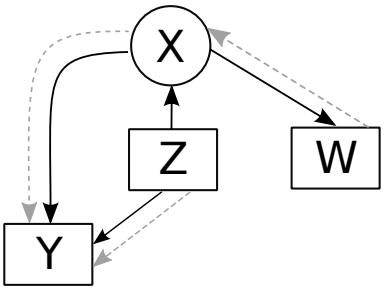

FIG 2. *Graphical structure of the error models. Rectangular boxes illustrate variables that are observed, while circles indicate unknown variables. Black arrows correspond to the classical, and the dashed grey arrows to the Berkson error model. The fundamental difference is the change in direction of the arrow between* $\boldsymbol{x}$ *and* $\boldsymbol{w}$.

response variable $\boldsymbol{y}$ is gained through $\boldsymbol{w}$ (Carroll et al., 2006). Ideally, repeated measurements of the true value x_i are available, so that

$$w_{ij} \,|\, x_i \sim \mathcal{N}(x_i, \tau_u) \ , \tag{2}$$

where w_{ij} is the j^{th} measurement of x_i. More generally, the error structure can be heteroscedastic with $\boldsymbol{w}_{\cdot j} \sim \mathcal{N}(\boldsymbol{x}, \tau_u \mathbf{D})$, where $\boldsymbol{w}_{\cdot j}$ denotes the vector of the j^{th} measurements, and the entries in the diagonal matrix $\mathbf{D}$ represent weights d_i that are proportional to the individual error precision $\tau_{u_i} := \tau_u(x_i)$ depending on x_i. In fact, both the homo- and heteroscedastic cases are relevant in practice (see, e.g., Subar et al. (2001) or example 5.1 presented here).

Parameter estimates belonging to $\boldsymbol{x}$ are usually attenuated in the classical ME setting. Consider for instance a simple linear regression with homoscedastic ME. Fitting the *naive* model $\boldsymbol{y} = \beta_0^\star + \beta_x^\star \boldsymbol{w} + \boldsymbol{\varepsilon}^\star$ instead of the true model $\boldsymbol{y} = \beta_0 + \beta_x \boldsymbol{x} + \boldsymbol{\varepsilon}$ will result in $|\beta_x^\star| < |\beta_x|$, if the error variance $1/\tau_u$ is larger than zero. We therefore call $\beta_x^\star$ the naive estimate. The left panel of Figure 3 illustrates this attenuation affect. Another important effect is the significant increase of the variability around the regression line.

3.3 Berkson ME model

Berkson-type error can be observed in experimental settings, where the value of a covariate may correspond to, e.g. a predefined fixed dose, temperature or time interval, but the true values $\boldsymbol{x}$ may deviate from these planned values $\boldsymbol{w}$ due to imprecision in the realization. The second setting where Berkson-type error occurs is in epidemiological or biological studies, where, e.g., averages of exposures in areas are assigned to individuals living or working close-by. Examples are the application of fixed doses of herbicides in bioassay experiments (Rudemo, Ruppert and Streibig, 1989) or the radiation epidemiology study described in Kerber et al. (1993) and Simon et al. (1995). Such circumstances led to the Berkson error model (Berkson, 1950)

$$\boldsymbol{x} = \boldsymbol{w} + \boldsymbol{u} \ ,$$

where $\boldsymbol{u}$ and $\boldsymbol{w}$ are independent, and

$$\boldsymbol{x} \,|\, \boldsymbol{w} \sim \mathcal{N}(\boldsymbol{w}, \tau_u \mathbf{D}) \ , \tag{3}$$

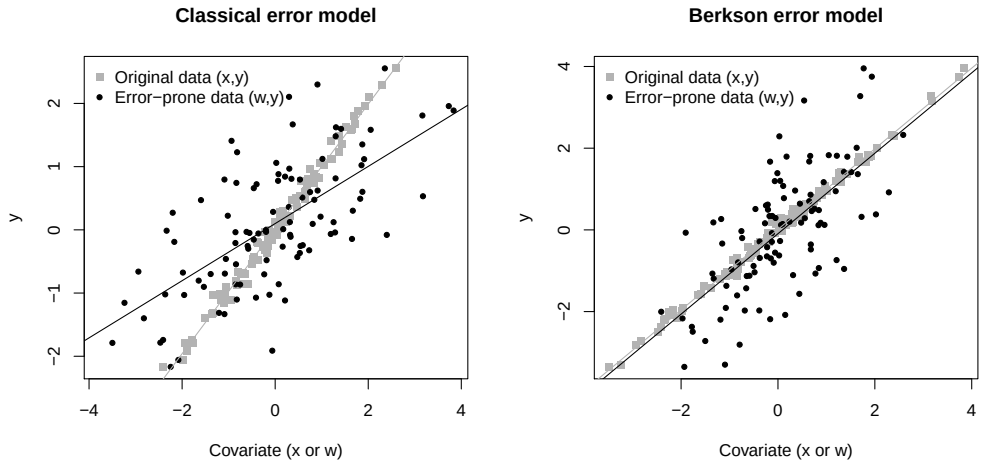

FIG 3. *Effect of ME in the linear model. Left: Classical ME. Two effects can be seen: 1) The absolute value of the covariate estimate is biased (attenuated); 2) The variability around the regression line in the data with ME (black circles) is much larger than in the case of the truly observed data (grey squares). Right: Berkson ME. The absolute value of the covariate estimate is unbiased in the linear model, while the variability around the regression line is larger for the data with ME.*

with $\mathbf{D}$ denoting a diagonal matrix as in Section 3.2. Like classical ME, the Berkson error is also assumed to be non-differential. The effect of Berkson error is fundamentally different from that of classical error. In the linear regression model there is no attenuation effect, as illustrated in the right panel of Figure 3. However, the residual precision suffers from the same qualitative bias as in the classical ME model. Issues become more involved for GL(M)Ms. For instance, parameter estimates for logistic regression are only approximately consistent in the Berkson case (Burr, 1988; Bateson and Wright, 2010), which makes error modeling essential.

The fundamental difference between the classical and Berkson error models implies different relationships between the error variances. Denote with τ_x^{-1} and τ_w^{-1} the variances of $\boldsymbol{x}$ and $\boldsymbol{w}$, respectively. Due to the independence assumption of $\boldsymbol{x}$ and $\boldsymbol{u}$ in the classical and between $\boldsymbol{w}$ and $\boldsymbol{u}$ in the Berkson error case, the variances in the two ME models can be written as

$$\begin{aligned} \tau_w^{-1} &= \tau_x^{-1} + \tau_u^{-1} \quad \text{(classical)} \ , \\ \tau_x^{-1} &= \tau_w^{-1} + \tau_u^{-1} \quad \text{(Berkson)} \ . \end{aligned}$$

Thus, the surrogate $\boldsymbol{w}$ is more variable than the true covariate $\boldsymbol{x}$ in the classical model, whereas the opposite is true in the Berkson case. This effect can also be observed in Figure 3.

4. MEASUREMENT ERROR MODELING WITH INLA

Bayesian analyses of ME problems date back to the seminal work of Clayton (1992) and are usually based on hierarchical models. Bayesian hierarchical models often require the use of MCMC algorithms, which typically are time-consuming and require diagnostic checks to ensure good mixing properties and convergence

of the simulated samples. These might be some of the reasons why Bayesian ME analysis is not widely used in practice.

4.1 Inference for latent Gaussian models based on deterministic approximations

Rue, Martino and Chopin (2009) proposed with INLA a new approach based on accurate approximations to perform fast Bayesian inference in a sub-class of hierarchical models, namely latent Gaussian models. The observation variable y_i is assumed to belong to a distribution family where the mean μ_i is linked to the linear predictor η_i via a link function $h^{-1}(.)$, so that $\eta_i = h^{-1}(\mu_i)$. The likelihood can be controlled by hyperparameters $\boldsymbol{\theta}_1$, e.g., in the Gaussian case the residual variance. To account for the effect of different covariates, the linear predictor is additively composed:

$$\eta_i = \beta_0 + \sum_{j=1}^{n_f} \omega_{ji} f^{(j)}(u_{ji}) + \boldsymbol{\beta}_z^\top \boldsymbol{z}_i + \varepsilon_i \ . \tag{4}$$

As in Equation (1b), $\boldsymbol{\beta}_z$ represents the vector of fixed effects of covariates $\boldsymbol{z}$. The $\{f^{(j)}\}$s are unknown functions of the covariates $\boldsymbol{u}$, where ω_{ij} denote known weights defined for each observed data point. The ε_is are additional unstructured terms.

A latent Gaussian model is obtained by assigning to $\boldsymbol{v} = (\boldsymbol{\eta}, \beta_0, \boldsymbol{f}, \boldsymbol{\beta}_z)$ a Gaussian prior with precision matrix $\mathbf{Q}(\boldsymbol{\theta}_2)$, controlled by hyperparameters $\boldsymbol{\theta}_2$. Of note, $\boldsymbol{v}$ is reparameterized to include $\boldsymbol{\eta}$ instead of $\boldsymbol{\varepsilon}$. The reason is that INLA (originally) requires that each data point y_i is linked to a single component of the latent field, namely η_i (Martino and Rue, 2010). Hence, a small random noise, ε_i, with high precision is always automatically added to the model. Finally, we need to specify a prior distribution on $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)$.

Assuming that y_i, $i = 1, \ldots, n$, are conditionally independent given $\boldsymbol{v}$ and $\boldsymbol{\theta}$, the posterior distribution is

$$\mathrm{p}(\boldsymbol{v}, \boldsymbol{\theta} \,|\, \boldsymbol{y}) \propto \mathrm{p}(\boldsymbol{\theta})\, \mathrm{p}(\boldsymbol{v} \,|\, \boldsymbol{\theta}) \prod_i \mathrm{p}(y_i \,|\, v_i, \boldsymbol{\theta}) \ ,$$

where $\mathrm{p}(\cdot|\cdot)$ denotes the conditional density of its arguments. The posterior marginal distributions of interest are

$$\mathrm{p}(v_i \,|\, \boldsymbol{y}) = \int_{\boldsymbol{\theta}} \mathrm{p}(v_i \,|\, \boldsymbol{\theta}, \boldsymbol{y})\, \mathrm{p}(\boldsymbol{\theta} \,|\, \boldsymbol{y}) d\boldsymbol{\theta} \ , \tag{5}$$

$$\mathrm{p}(\theta_j \,|\, \boldsymbol{y}) = \int_{\boldsymbol{\theta}_{-j}} \mathrm{p}(\boldsymbol{\theta} \,|\, \boldsymbol{y})\ d\boldsymbol{\theta}_{-j} \ . \tag{6}$$

Except for cases when everything can be computed analytically, exact inference is very challenging. Hence, sampling-based approaches have been the standard tool (Gelfand and Smith, 1990). Currently, few generic software packages based on MCMC, e.g. OpenBugs (Lunn et al., 2009), are available.

INLA avoids sampling and instead provides accurate approximations to (5) and (6). First, it builds a Laplace approximation to $\mathrm{p}(\boldsymbol{\theta} \,|\, \boldsymbol{y})$, from which a representative set of support points $\{\boldsymbol{\theta}_k\}$ is selected. For each $\boldsymbol{\theta}_k$ a Laplace approximation (or its simplified version) is built to $\mathrm{p}(v_i \,|\, \boldsymbol{\theta}_k, \boldsymbol{y})$. Finally, to obtain approximations

of $\mathrm{p}(v_i \mid \boldsymbol{y})$, the integral in Equation (5) is numerically computed as a finite sum. Numerical approximations of $\mathrm{p}(\theta_j \mid \boldsymbol{y})$ are obtained similarly.

As discussed in Rue, Martino and Chopin (2009) and illustrated in a variety of different applications, the approximation error of INLA is small compared to the Monte Carlo error and negligible in practice, see for example Paul et al. (2010); Schrödle et al. (2011); Riebler, Held and Rue (2012). For details on the approximations we refer to Rue, Martino and Chopin (2009).

4.2 Classical ME model inference

Consider a generalized linear (mixed) model regressing a response $\boldsymbol{y}$ on covariates $\boldsymbol{x}$ and $\boldsymbol{z}$. The covariate $\boldsymbol{z}$ can be observed directly, while instead of $\boldsymbol{x}$ only a surrogate $\boldsymbol{w}|\boldsymbol{x}, \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{x}, \tau_u \mathbf{D})$, following the classical error model (2), is available. This model is not covered by the general INLA model set (4). The problem is that it includes a product structure of two unknown components, namely β_x and $\boldsymbol{x}$. Here, β_x is no longer a fixed effect, but plays the role of an unknown scaling factor (hyperparameter) for the latent process $\boldsymbol{x}$.

To fit within the scope of INLA, $\boldsymbol{x}$ must follow a Gaussian model. As the intercept β_0, covariates $\boldsymbol{z}$ and the unstructured terms ε_i do not influence the calculations, let us consider a simplified model, where only the product term of interest remains:

$$\boldsymbol{\eta} = \beta_x \boldsymbol{x} \ , \tag{7}$$

$$\begin{aligned} \boldsymbol{w} &= \boldsymbol{x} + \boldsymbol{u} \ , \\ \boldsymbol{x} &= \alpha_0 + \boldsymbol{\mu} \ , \end{aligned} \tag{8}$$

with $\boldsymbol{u} \sim \mathcal{N}(0, \tau_u \mathbf{D})$ and $\boldsymbol{\mu} \sim \mathcal{N}(0, \tau_x \mathbf{I})$. Further, let $\alpha_0 \sim \mathcal{N}(0, 10^{-4})$, $\beta_x \sim \mathcal{N}(0, 10^{-4})$, $\tau_x \sim \mathrm{G}(a_x, b_x)$, and $\tau_u \sim \mathrm{G}(a_u, b_u)$. Other priors for the precisions τ_x and τ_u could also be used in INLA, see Roos and Held (2011) for an example. Set $\boldsymbol{\theta} = (\beta_x, \tau_x, \tau_u, \alpha_0)$. The latent field in this simplified model contains only $\boldsymbol{x}$. Hence, the posterior distribution of $\boldsymbol{x}$ and $\boldsymbol{\theta}$ is

$$\mathrm{p}(\boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{w}) \ \propto \ \mathrm{p}(\boldsymbol{\theta}) \ \mathrm{p}(\boldsymbol{x} \mid \boldsymbol{\theta}) \ \mathrm{p}(\boldsymbol{w} \mid \boldsymbol{x}, \boldsymbol{\theta}) \ \mathrm{p}(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) \ .$$

Using $\mathrm{p}(\boldsymbol{x} \mid \boldsymbol{\theta}) \, \mathrm{p}(\boldsymbol{w} \mid \boldsymbol{x}, \boldsymbol{\theta}) = \mathrm{p}(\boldsymbol{x} \mid \boldsymbol{w}, \boldsymbol{\theta}) \, \mathrm{p}(\boldsymbol{w} \mid \boldsymbol{\theta})$, we get

$$\mathrm{p}(\boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{w}) \ \propto \ \mathrm{p}(\boldsymbol{\theta}) \ \mathrm{p}(\boldsymbol{x} \mid \boldsymbol{w}, \boldsymbol{\theta}) \ \mathrm{p}(\boldsymbol{w} \mid \boldsymbol{\theta}) \ \mathrm{p}(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) \ , \tag{9}$$

where

$$\boldsymbol{w} \mid \boldsymbol{\theta} \ \sim \ \mathcal{N}\left(\alpha_0 \mathbf{1}, \left[(\tau_u \mathbf{D})^{-1} + (\tau_x \mathbf{I})^{-1}\right]^{-1}\right) \ .$$

Since $\boldsymbol{x}$ only enters in one term in (9) (apart from the likelihood), it can be treated as an ordinary latent model, where $\boldsymbol{x}|\boldsymbol{w}, \boldsymbol{\theta}$ is Gaussian:

$$\begin{aligned} \mathrm{p}(\boldsymbol{x} \mid \boldsymbol{w}, \boldsymbol{\theta}) \ &\propto \ \mathrm{p}(\boldsymbol{x} \mid \boldsymbol{\theta}) \, \mathrm{p}(\boldsymbol{w} \mid \boldsymbol{x}, \boldsymbol{\theta}) \\ &\propto \ \exp\left(-\frac{\tau_x}{2}(\boldsymbol{x} - \alpha_0 \mathbf{1})^\top (\boldsymbol{x} - \alpha_0 \mathbf{1}) - \frac{\tau_u}{2}(\boldsymbol{x} - \boldsymbol{w})^\top \mathbf{D} (\boldsymbol{x} - \boldsymbol{w})\right) \ . \end{aligned}$$

Combining the quadratic forms gives

$$\boldsymbol{x} \mid \boldsymbol{w}, \boldsymbol{\theta} \ \sim \ \mathcal{N}\left((\tau_x \alpha_0 \mathbf{1} + \tau_u \mathbf{D} \boldsymbol{w})(\tau_x \mathbf{I} + \tau_u \mathbf{D})^{-1} \ , \ \tau_x \mathbf{I} + \tau_u \mathbf{D}\right) .$$

To arrive at a more convenient model formulation, we apply a change of variables and set

$$\beta_x \boldsymbol{x} \to \boldsymbol{\nu} \ ,$$

so that $\boldsymbol{\eta} = \boldsymbol{\nu}$. Following the same steps as above but replacing $\boldsymbol{x}$ with $\boldsymbol{\nu}$, we can derive a latent model for $\boldsymbol{\nu}|\boldsymbol{w}, \boldsymbol{\theta}$:

$$\boldsymbol{\nu} \,|\, \boldsymbol{w}, \boldsymbol{\theta} \ \sim \ \mathcal{N}\left(\beta_x(\tau_x\alpha_0\mathbf{1} + \tau_u\mathbf{D}\boldsymbol{w})(\tau_x\mathbf{I} + \tau_u\mathbf{D})^{-1}, \ \frac{\tau_x\mathbf{I} + \tau_u\mathbf{D}}{\beta_x^2}\right) \ .$$

This model is termed "`mec`" within the `R`-package `r-inla` and has four hyperparameters: β_x, τ_x, τ_u, α_0.

The `mec` model could be extended to a more complex setting, where the covariate $\boldsymbol{x}$ is regarded as a random variable depending on other covariates, leading to a *structural model* with prior distribution

$$\boldsymbol{x} \,|\, \boldsymbol{z} \sim \mathcal{N}(\alpha_0 + \boldsymbol{z}\boldsymbol{\alpha}_z, \tau_x\mathbf{I}) \ . \tag{10}$$

Here, α_0 is the intercept, $\boldsymbol{\alpha}_z$ is the $p \times 1$ vector of fixed effects, and τ_x^{-1} the residual variance in the linear regression of $\boldsymbol{x}$ on $\boldsymbol{z}$. Equation (10) is the *exposure model* and it is often used in epidemiological studies (Gustafson, 2004). The assumption that the distribution of the unobserved $\boldsymbol{x}$ given the observable covariates $\boldsymbol{z}$ follows a normal distribution is crucial to apply INLA, but often justified. Due to recent extensions of INLA, see Martins and Rue (2012), which allow that independent components of the latent field can have a non-Gaussian distribution, this assumption might be relaxed in the future.

For exposure models including many components, the model specification with `mec` becomes complex. Recently, a new copy feature was added to the INLA package to treat situations where elements of a latent field $\boldsymbol{x}$ are needed more than once for each observation, consider for example bivariate normally distributed random effects (Martins et al., 2012). An almost identical copy denoted by $\boldsymbol{x}^\star$ for $\boldsymbol{x}$ is then created and the latent field is extended to $\boldsymbol{x}_c = (\boldsymbol{x}, \boldsymbol{x}^\star)$ with $\pi(\boldsymbol{x}_c) = \mathrm{p}(\boldsymbol{x})\,\mathrm{p}(\boldsymbol{x}^\star \,|\, \boldsymbol{x})$ and

$$\mathrm{p}(\boldsymbol{x}^\star \,|\, \boldsymbol{x}, \psi, \tau) \propto \exp\left(-\frac{\tau}{2}(\boldsymbol{x}^\star - \psi\boldsymbol{x})^\top(\boldsymbol{x}^\star - \psi\boldsymbol{x})\right).$$

The precision τ, fixed to some large value, controls the similarity between $\boldsymbol{x}^\star$ and $\boldsymbol{x}$ (default value: 10^9). A copied model may contain an unknown scale parameter ψ, which per default is fixed to one. Our framework can be considered as such a copy setting, where $\boldsymbol{x}$ appears in different levels of the model (compare the simplified model (7), (8), (10)), instead of repeated times for one observation. The parameter β_x from (7) can thereby be regarded as the unknown scaling factor ψ of the copied model. ME models with general exposure structures as in (10) are hence straightforward to analyze with INLA, compare application 5.2 and the corresponding Supplementary Material.

The general model structure including (1b), (2) and (10) is hierarchical and can be written as

$$\begin{aligned} E(\boldsymbol{y}) &= h(\beta_0 + \beta_x\boldsymbol{x} + \boldsymbol{z}\boldsymbol{\beta}_z) \ , \\ \mathbf{0} &= -\boldsymbol{x} + \alpha_0 + \boldsymbol{z}\boldsymbol{\alpha}_z + \boldsymbol{\varepsilon}_x \ , \quad \boldsymbol{\varepsilon}_x \sim \mathcal{N}(0, \tau_x\mathbf{I}) \ , \\ \boldsymbol{w} &= \boldsymbol{x} + \boldsymbol{u} \ , \qquad \boldsymbol{u} \sim \mathcal{N}(0, \tau_u\mathbf{D}) \ . \end{aligned} \tag{11}$$

Note that the exposure model (10), encoded in (11), can be easily extended to include structured or unstructured random effects terms. Implementation in INLA requires a *joint model* formulation, where the column of the actual response $\boldsymbol{y}$ is merged with a column of pseudo-observations $\mathbf{0}$, compare (11), to account for the exposure model (see also Schrödle, Held and Rue (2012) and Ruiz-Cárdenas, Krainski and Rue (2012)), and a column of the observed values $\boldsymbol{w}$ to model the error. The resulting INLA response matrix contains one separate column per equation, namely

$$\begin{bmatrix} y_1 & \texttt{NA} & \texttt{NA} \\ \vdots & \vdots & \vdots \\ y_n & \texttt{NA} & \texttt{NA} \\ \texttt{NA} & 0 & \texttt{NA} \\ \vdots & \vdots & \vdots \\ \texttt{NA} & 0 & \texttt{NA} \\ \texttt{NA} & \texttt{NA} & \overline{w}_{1\cdot} \\ \vdots & \vdots & \vdots \\ \texttt{NA} & \texttt{NA} & \overline{w}_{n\cdot} \end{bmatrix} .$$

Here, $\overline{w}_{i\cdot}$ denotes the mean from the repeated measurements w_{ij} of x_i, which is commonly used as proxy for the latter. Each observation column requires the assignment of a different likelihood function. The latter two are always assumed to be Gaussian (or near-Gaussian), while the first follows the specified exponential family distribution. In most software packages the assignment of various likelihood functions is not straightforward, if possible at all. In INLA the only requirement is that the response observations, given the latent field and the hyperparameters, are conditionally independent.

4.3 Berkson ME model inference

Consider again a generalized linear (mixed) model regression (1b), but replace the classical error model (2) by the Berkson model (3)

$$\boldsymbol{x} \,|\, \boldsymbol{w}, \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{w}, \tau_u \mathbf{D}) \ .$$

Since $\boldsymbol{x}$ is defined conditionally on the observations $\boldsymbol{w}$, the exposure model (10) becomes obsolete. The same simplifications as in (7) lead to the hierarchical model

$$\begin{aligned} \boldsymbol{\eta} &= \beta_x \boldsymbol{x} \ , \\ \boldsymbol{x} &= \boldsymbol{w} + \boldsymbol{u} \ \ , \end{aligned} \tag{12}$$

with $\boldsymbol{u} \sim \mathcal{N}(0, \tau_u \mathbf{D})$, $\beta_x \sim \mathcal{N}(0, 10^{-4})$ and $\tau_u \sim \mathrm{G}(a_u, b_u)$. The hyperparameters are $\boldsymbol{\theta} = (\beta_x, \tau_u)$. Importantly, the latent model $\boldsymbol{x}|\boldsymbol{w}, \boldsymbol{\theta}$ is now identical to the error model (3), because the latent field contains only $\boldsymbol{x}$. It is thus straightforward to calculate the posterior distribution

$$\mathrm{p}(\boldsymbol{x}, \boldsymbol{\theta} \,|\, \boldsymbol{y}, \boldsymbol{w}) \ \propto \ \mathrm{p}(\boldsymbol{\theta}) \ \mathrm{p}(\boldsymbol{x} \,|\, \boldsymbol{w}, \boldsymbol{\theta}) \ \mathrm{p}(\boldsymbol{y} \,|\, \boldsymbol{x}, \boldsymbol{\theta}) \ .$$

The same change of variables $\beta_x \boldsymbol{x} \to \boldsymbol{\nu}$ as in the classical case is useful again, and leads to

$$\boldsymbol{\nu} \,|\, \boldsymbol{w}, \boldsymbol{\theta} \ \sim \ \mathcal{N}\left(\beta_x \boldsymbol{w}, \ \frac{\tau_u}{\beta_x^2} \mathbf{D}\right) \ .$$

This model is termed "meb" within the R-package r-inla and has two hyperparameters: β_x and τ_u.

As in the classical case, the copy feature might be used alternatively, see Section 4.2. However, here it does not add to the generality of the model specification as no exposure model is involved. Thus, we recommend the use of the more straightforward model definition using the meb model, and just illustrate for completeness the formulation using the copy feature. Since the respective joint model contains only the two components

$$\begin{aligned} E(\boldsymbol{y}) &= h(\beta_0 + \beta_x \boldsymbol{x} + \boldsymbol{z}\boldsymbol{\beta}_z) \ , \\ -\boldsymbol{w} &= -\boldsymbol{x} + \boldsymbol{u} \ , \end{aligned}$$

the response matrix simplifies to

$$\begin{bmatrix} y_1 & \texttt{NA} \\ \vdots & \vdots \\ y_n & \texttt{NA} \\ \texttt{NA} & -w_1 \\ \vdots & \vdots \\ \texttt{NA} & -w_n \end{bmatrix} .$$

The latent field is given by $\boldsymbol{v} = (\boldsymbol{x}^\top, \beta_0, \boldsymbol{\beta}_z^\top)^\top$, and $\boldsymbol{\theta} = (\beta_x, \tau_u, \boldsymbol{\theta}_1^\top)^\top$ are the hyperparameters, where $\boldsymbol{\theta}_1$ may again contain additional hyperparameters of the likelihood. All components in $\boldsymbol{v}$ and the coefficient β_x obtain Gaussian priors, and the error precision τ_u a suitable gamma prior.

5. APPLICATIONS

In the following, we demonstrate how to define the different measurement error applications introduced in Section 2 in the INLA framework. The respective r-inla code is given in the Supplementary Material. A comparison of the results obtained by INLA to those obtained by an independent MCMC implementation is provided for each application to highlight the accuracy of INLA.

Note that the efficiency of sampling-based techniques, such as MCMC, is generally improved when the covariates are centered around zero (Gelfand, Sahu and Carlin, 1995, 1996). The same is true for INLA. Here, covariates are sometimes not centered to compare to the results of a reference example. Additional adjustments of the default parameters in the numerical optimization routine of INLA might be needed then, see Supplementary Material.

5.1 Inbreeding in Swiss ibex populations

The ibex data introduced in Section 2.1 were analyzed using a linear model with classical heteroscedastic error variances. The observation model is a simple Gaussian response model

$$\boldsymbol{y} \,|\, \boldsymbol{x} \sim \mathcal{N}(\beta_0 + \beta_x \boldsymbol{x}, \tau_\varepsilon \mathbf{I})$$

with $\boldsymbol{y}$ being the intrinsic growth rates and $\boldsymbol{x}$ the inbreeding coefficients of the populations. The level of inbreeding x_i in population $i = 1, \ldots, 26$ was estimated as w_i from a Bayesian analysis, which, as a by-product, also provided an

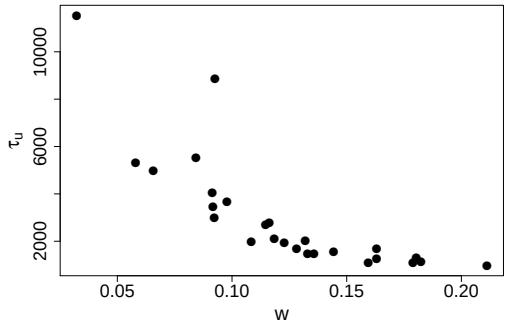

FIG 4. *Uncertainty in the covariate* $\boldsymbol{x}$ *in the ibex study, depending on the estimate* $\boldsymbol{w}$. *Larger values can be estimated with less precision (i.e., larger variance* $1/\tau_u$*).*

estimated population-specific error precision $\hat{\tau}_{u_i}$. Larger values of $\boldsymbol{w}$ have more uncertainty, i.e., smaller precision, as shown in Figure 4. It is natural to formulate a heteroscedastic classical error model

$$\boldsymbol{w} \,|\, \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}, \tau_u \mathbf{D})$$

with entries $\hat{\tau}_{u_i}$ in the diagonal matrix $\mathbf{D}$. Since $\boldsymbol{x}$ is the only covariate in the model, the exposure model (10) reduces to

$$\boldsymbol{x} \sim \mathcal{N}(\alpha_0 \mathbf{1}, \tau_x \mathbf{I}) \ .$$

Here, $\alpha_0 = 0$ was fixed due to the preceding covariate centering. The unknowns in this example are the latent field $\boldsymbol{v} = (\boldsymbol{x}^\top, \beta_0)^\top$ and the hyperparameters $\boldsymbol{\theta} = (\beta_x, \tau_u, \tau_x, \tau_\varepsilon)^\top$. We assigned independent $\mathcal{N}(0, 10^{-4})$ priors to β_0 and β_x. All precisions were given gamma priors with mode fixed at some estimated value. This led to $\tau_u \sim \mathrm{G}(2, 1)$ with mode equal to one, $\tau_\varepsilon \sim \mathrm{G}(10, 0.387)$ with mode at half of the residual precision of the regression $\boldsymbol{y} \sim \boldsymbol{w}$, and $\tau_x \sim \mathrm{G}(10, 0.0117)$ with mode at $\hat{\tau}_x = [1/\hat{\tau}_w - 1/\hat{\tau}_u]^{-1}$, where $\hat{\tau}_w$ is the sampling precision of $\boldsymbol{w}$, and $\hat{\tau}_u$ is the mean of all estimates $\hat{\tau}_{u_i}$. The use of informative priors is essential here, because the model is not identifiable if the error precision τ_u is not restricted.

An MCMC simulation was run for 100'000 iterations with a burn-in of 10'000 iterations and a saving frequency of 10. The estimates obtained from INLA were chosen as starting values. Convergence was visually checked. Figure 5 shows the perfect fit between the MCMC samples and the posterior marginals of INLA. Of note, due to the Gaussian likelihood the results obtained by INLA are exact and contain no approximation error. The parameter estimates are graphically compared to the naive analysis, including $\boldsymbol{w}$ instead of $\boldsymbol{x}$, in Figure 6. The absolute value of the slope $|\beta_x|$ and the residual precision τ_ε are underestimated in the naive regression, as predicted by the theory. For further comparison, the results of an analysis with SIMEX are given, suggesting a slightly stronger correction for the slope β_x. SIMEX was run as described in Cook and Stefanski (1994) and Devanarayan and Stefanski (2002), using a quadratic extrapolation function and a bootstrap of 10'000 iterations to estimate the quantile confidence intervals. There also exists an `R`-package `simex` (Lederer and Küchenhoff, 2009), but it did not allow for heteroscedastic error variances at the time when this analysis was carried out.

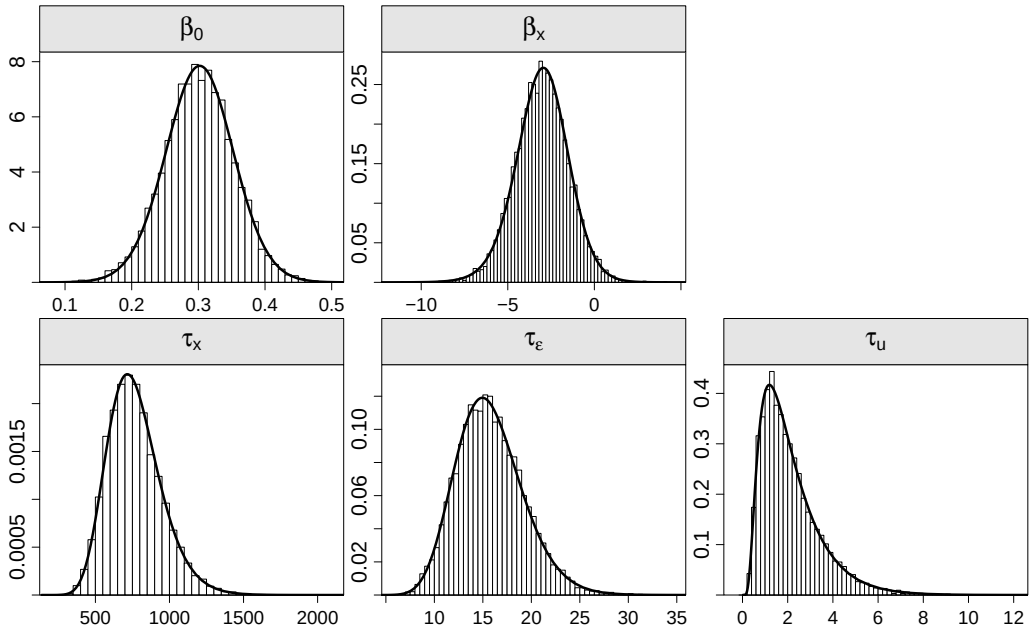

FIG 5. *Comparison of the MCMC samples (histograms) with the INLA posterior marginals (lines) for the ibex data.*

As was mentioned in Section 2.1 and is obvious from Figure 6, the difference between the naive and the corrected estimate of β_x is not extreme (about 17%). Yet, in this example the difference is just enough to remove 0 from the confidence interval.

5.2 Influence of systolic blood pressure on coronary heart disease

Since SBP has been measured twice at different examinations, the magnitude of the measurement error can be quantified. Here, we assume that the average values $\boldsymbol{w}$ between $\boldsymbol{w_1} = \log(\text{SBP}_1 - 50)$ at first examination and $\boldsymbol{w}_2 = \log(\text{SBP}_2 - 50)$ at second examination are independent and normally distributed with mean $\boldsymbol{x}$ and precision τ_u, leading to the classical homoscedastic error model

$$\boldsymbol{w} \,|\, \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}, \tau_u \mathbf{I}) \ .$$

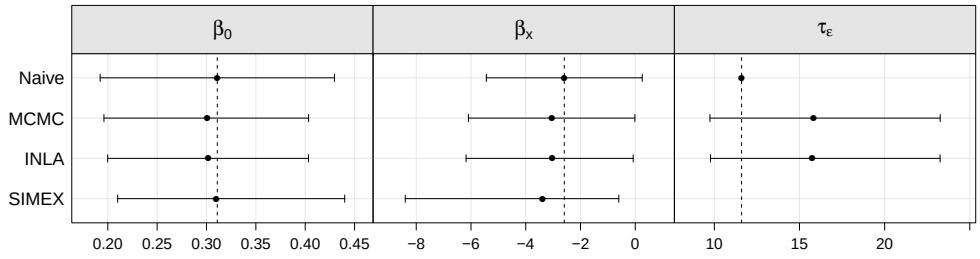

FIG 6. *Parameter estimates and 95% confidence/credible intervals in the ibex data analysis. The dashed lines indicate the naive estimates. Only parameters where a naive estimate was available are included here.*

The observation model is logistic, incorporating smoking status $\boldsymbol{z}$ and long-term blood pressure $\boldsymbol{x}$ as covariates:

$$\boldsymbol{\eta} = \text{logit}\left[P(\boldsymbol{y} = 1 \,|\, \boldsymbol{x}, \boldsymbol{z})\right] = \beta_0 + \beta_x \boldsymbol{x} + \beta_z \boldsymbol{z} \ .$$

Finally, the exposure model (10) comes in its most general form

$$\boldsymbol{x} \,|\, \boldsymbol{z} \sim \mathcal{N}(\alpha_0 + \alpha_z \boldsymbol{z}, \tau_x \mathbf{I}) \ .$$

The unknowns in this model are $\boldsymbol{v} = (\boldsymbol{x}^\top, \beta_0, \beta_z, \alpha_0, \alpha_z)^\top$ and $\boldsymbol{\theta} = (\beta_x, \tau_u, \tau_x)^\top$. The α- and β-parameters were given independent $\mathcal{N}(0, 10^{-4})$ priors. The prior for τ_u was selected by using the repeated measurements $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ as follows: starting from an improper reference gamma prior $\text{G}(a_0, b_0) = \text{G}(0, 0)$ for τ_u, the parameters of the final gamma prior $\text{G}(a_u, b_u)$ are deduced as

$$\begin{aligned} p(\tau_u \,|\, \boldsymbol{w}_1, \boldsymbol{w}_2) &\propto p(\boldsymbol{w}_1, \boldsymbol{w}_2 \,|\, \tau_u) \cdot p(\tau_u) \\ &\propto \tau_u^{a_0 - 1 + n} \cdot \exp\left(-\frac{\tau_u}{2} \sum_i [(w_{i1} - \overline{w}_{i\cdot})^2 + (w_{i2} - \overline{w}_{i\cdot})^2] - \tau_u \cdot b_0 \right) \\ &\sim \text{G}(n, \frac{1}{2} \sum_i [(w_{i1} - \overline{w}_{i\cdot})^2 + (w_{i2} - \overline{w}_{i\cdot})^2]) \ , \end{aligned} \tag{13}$$

where $\overline{w}_{i\cdot} = (w_{i1} + w_{i2})/2$ is the i^{th} entry of the vector $\boldsymbol{w}$. Note that (13) is the prior precision for $\boldsymbol{w}$ (not for the repeated measurements $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$). Following Example 5.1, τ_x was given a gamma prior $\text{G}(100, 4.02)$ with mode at the estimated value $\hat{\tau}_x = \left[1/\hat{\tau}_{w|z} - 1/\hat{\tau}_u\right]^{-1}$, where $\hat{\tau}_{w|z}$ is the residual precision of $\boldsymbol{w}$ given $\boldsymbol{z}$. Again, the use of informative priors is essential to avoid identifiability problems. The prior specifications might deviate from the reference example in Carroll et al. (2006), where the exact parameters were not given, but it is important to use the knowledge from repeated measurements and observed values in order to obtain sensible results. Furthermore, note that Carroll et al. (2006) used the quantity $\Delta := \tau_x/\tau_u$ instead of τ_u, and gave it a uniform prior in the interval $(0, 0.5)$. Since this is not straightforward with INLA, the model was modified as described.

To obtain Markov Chain Monte Carlo (MCMC) posterior marginals, regression coefficients of GLMs cannot directly be sampled from a standard full conditional distribution. Here, samples were obtained according to Gamerman (1997). The algorithm can be used if the observations y_i are conditionally independent and follow an exponential family density. For the regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_x, \boldsymbol{\beta}_z^\top)^\top$, this approach uses transition densities that combine the weighted least squares method with a prior on $\boldsymbol{\beta}$ (McCullagh and Nelder, 1989; West, 1985). The full conditionals for the unknowns in our logistic regression model are given in the Supplementary Material.

The simulation was run for 200'000 iterations with a burn-in of 10'000 and every 10^{th} value was saved. Starting values for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ were chosen from the INLA output. For τ_u and τ_x, the initial estimates $\hat{\tau}_u = n/\sum_{i=1}^n \left[(w_{i,1} - \overline{w}_{i\cdot})^2 + (w_{i,2} - \overline{w}_{i\cdot})^2\right]$ and $\hat{\tau}_x$ from above were used. The acceptance rates for $\boldsymbol{\beta}$ and $\boldsymbol{x}$ were 83% and 69%, respectively.

The agreement between the MCMC and INLA output is almost perfect. The corresponding Figure containing the MCMC histograms and the INLA posterior marginals is included in the Supplementary Material. Figure 7 shows parameter

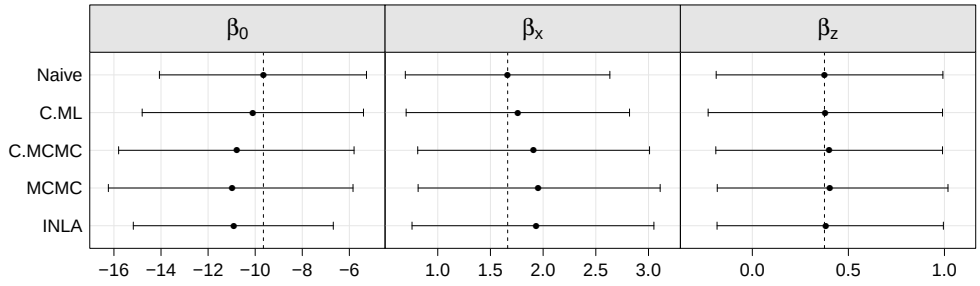

FIG 7. *Parameter estimates and 95% confidence/credible intervals for the Framingham data analysis. For MCMC and INLA, posterior means are used as point estimates.* C.MCMC *and* C.ML *stand for the Bayesian and the maximum likelihood analysis conducted in Carroll et al. (2006). The dashed lines indicate the naive estimates.*

estimates for the naive regression model including $\boldsymbol{w}$, and for four error-correction approaches. Note that only parameters where the naive estimate was available were included. Carroll et al. (2006) analyzed the data with a maximum-likelihood method and a Bayesian approach using MCMC. The fourth and fifth rows show the results as obtained from MCMC and INLA with the setup described here. All error-corrected estimates and the confidence/credible intervals are similar.

5.3 Seedling growth across different light conditions

Let $\boldsymbol{y}$ denote the number of new leaves per plant after a four months growth phase. The covariate $\boldsymbol{z}$ denotes the degree of defoliation and $\boldsymbol{x} = \log(\%\text{light})$ the light intensity, where $\boldsymbol{w}$ is the respective target value. Using $\boldsymbol{w}$ instead of $\boldsymbol{x}$ in the analysis leads to the homoscedastic Berkson error model

$$\boldsymbol{x} \,|\, \boldsymbol{w} \sim \mathcal{N}(\boldsymbol{w}, \tau_u \mathbf{I}) \ .$$

In the following we centered both covariates $\boldsymbol{w}$ and $\boldsymbol{z}$. This data structure leads to a Poisson regression model with nested design. To account for overdispersion, independent normal random effects $\gamma_{ijk} \sim \mathcal{N}(0, \tau_\gamma)$ have been added, extending the GLM to a GLMM:

$$\eta_{ijk} = \log(E[y_{ijk} \,|\, \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{\gamma}]) = \beta_0 + \beta_x x_{ij} + \beta_z z_k + \gamma_{ijk} \ , \tag{14}$$

with $i = 1, 2, 3$ denoting the light condition, $j = 1, \ldots, 5$ the shadehouse per light condition, and $k = 1, \ldots, 4$ the degrees of defoliation. The unknowns of this model are $\boldsymbol{v} = (\boldsymbol{x}^\top, \beta_0, \beta_z)^\top$ and $\boldsymbol{\theta} = (\beta_x, \tau_u, \tau_\gamma)^\top$.

Priors were chosen as $\boldsymbol{\beta} \sim \mathcal{N}(0, 10^{-4})$, $\tau_u \sim \text{G}(100, 19.88)$ and $\tau_\gamma \sim \text{G}(100, 1.712)$, where the mode for the τ_u prior was fixed at the estimated value $\hat{\tau}_u = 14/\sum_{ij}(x_{ij} - w_{ij})^2 = 4.98$ from the individual shadehouse measurements $\boldsymbol{x}$ when compared to $\boldsymbol{w}$, and the mode for the τ_γ prior at the $\hat{\tau}_\gamma = 57.83$ estimate from the regression (14) including $\boldsymbol{w}$ instead of $\boldsymbol{x}$.

The results from the regression with INLA were compared to an MCMC run with 200'000 iterations and a burn-in of 10'000 iterations. Sampling was based on a reparameterization as proposed by Besag et al. (1995), where all except one full conditional distribution are standard and can be Gibbs-sampled. Figure 8, left, shows the posterior marginal for the coefficient of interest, β_x, as obtained

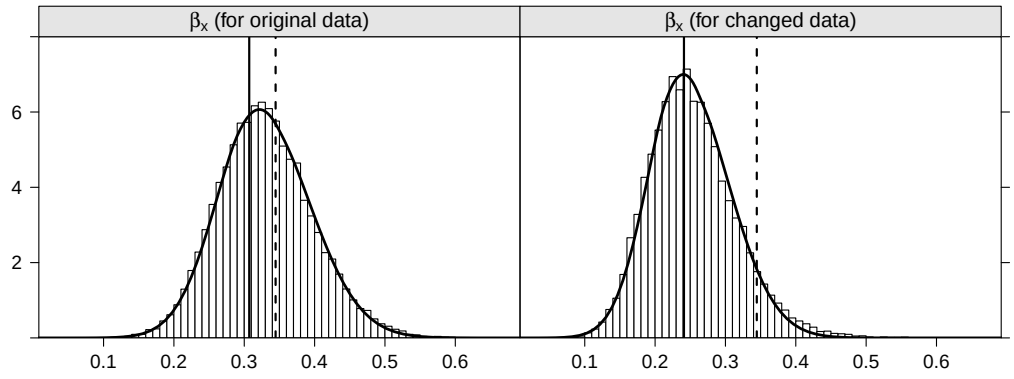

FIG 8. *Posterior marginals for MCMC (histograms) and INLA (lines) for the seedling growth example. The straight solid lines indicate the correct estimates of the slope β_x, while dashed lines mark the naive estimates when $\boldsymbol{w}$ is included in the regression. (Left) Original data. (Right) Data with five times larger error variance.*

by MCMC (histograms) and INLA (line). The solid and dashed lines mark the estimates when $\boldsymbol{x}$ or $\boldsymbol{w}$ are included in the regression, respectively. As can be seen, the bias due to the Berkson error is relatively small ($\hat{\beta}_x = 0.307$ vs. 0.345). To illustrate the size of a potential error correction, we now assume that the error variance $1/\tau_u$ was five times larger, i.e., the true values $\boldsymbol{x}$ were scattered broader around the target values $\boldsymbol{w}$. To this end, error was added artificially to $\boldsymbol{x}$. The analysis was, apart from that, repeated with the same data and the same model (including, of course, an adjusted prior for τ_u). The right panel of Figure 8 illustrates the much stronger bias induced by the increased Berkson error ($\hat{\beta}_x = 0.241$ vs. 0.345), and shows that the model approximately recovers the correct value. Marginals for the other regression parameters are not shown, as no noticeable bias was present.

Theory predicts that slope parameters in log-linear Poisson models do not suffer from a bias in the presence of Berkson ME (Carroll, 1989). The model used here nevertheless shows this bias (and it can be artificially enhanced), which might be due to small deviations from the ideal Poisson model assumptions and overdispersion.

6. DISCUSSION

Measurement error in covariates may lead to serious biases in parameter estimates and confidence intervals of statistical models, if it is not accounted for in the analysis. A variety of approaches to model such error have been proposed in the past decades, among which Bayesian methods probably provide the most flexible framework. Bayesian treatments, employing MCMC samplers, have been successfully applied for more than 20 years, but their application has never become part of standard regression analyses.

The aim of this work was to illustrate how the most common ME models (classical and Berkson error) can be included in GLMMs using the recently proposed INLA framework, which gives fast and accurate approximations instead of doing any sampling. The provided R-code should help to make such models accessible

to a broader audience. Note that INLA includes a much larger variety of likelihood functions and latent models than we could illustrate here, and the modular structure adds to its flexibility. It is, for instance, straightforward to treat several mismeasured covariates jointly, to introduce a systematic bias into the error model, or to include any structured random term into the model formulation.

One of the biggest challenges when treating mismeasured variables is the estimation of the error variance, either from repeated measurements, instrumental variables or from previous studies. The advantage of a Bayesian approach, as the one taken here, is that uncertainty of such estimates can be incorporated into prior distributions, while frequentist approaches require to fix such parameters. Moreover, sensitivity to chosen prior assumptions can be easily checked due to the computational speed of INLA, see Roos and Held (2011).

Gaussian classical and Berkson error models naturally fit into the INLA framework of latent Gaussian models, and thus the error-prone covariates used here are always continuous. The computational speed of INLA may allow to treat even misclassification error by following the misclassification SIMEX (MC-SIMEX) idea presented in Küchenhoff, Mwalili and Lesaffre (2006). Here, INLA would be applied repeatedly to estimate posterior marginals for different degrees of misclassification and extrapolating back to no misclassification. Such ideas need further investigation and should be detailed in the future.

SUPPLEMENTARY MATERIAL FOR "MEASUREMENT ERROR IN GLMMS WITH INLA"

Due to space constraints, the R-code for all examples presented here is described in detail in the supplementary document. Furthermore, this document contains full conditionals and posterior marginals for Section 5.2. On `www.r-inla.org/examples/case-studies/muff-etal-2013` selected data and R-code are provided for download.

ACKNOWLEDGEMENTS

We thank Prof. Lukas Keller for stimulating the ME project and many interesting discussions. We are grateful to Prof. Lukas Keller, Dr. Iris Biebach and Claudio Bozzuto for providing us with the ibex data. We also thank Prof. Helmut Küchenhoff for useful comments and suggestions to this work, and Dr. Malgorzata Roos for comments to the manuscript.

REFERENCES

Bateson, T. F. and Wright, J. M. (2010). Regression calibration for classical exposure measurement error in environmental epidemiology studies using multiple local surrogate exposures. *American Journal of Epidemiology* **172** 344–352.

Berkson, J. (1950). Are there two regressions? *Journal of the American Statistical Association* **45** 164-180.

Bernardinelli, L., Pasctto, C., Best, N. G. and Gilks, W. R. (1997). Disease mapping with errors in covariates. *Statistics in Medicine* **16** 741–752.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science* **10** 3–66.

Bozzuto, C., Biebach, I., Muff, S. and Keller, L. F. (2013). Inbreeding effects on population dynamics – the case of Swiss Ibex populations Technical Report, Institute of Evolutionary Biology and Environmental Studies, University of Zurich.

BUONACCORSI, J. P. (2010). *Measurement error. Interdisciplinary Statistics*. CRC Press, Boca Raton, FL. Models, methods, and applications.

BURR, D. (1988). On errors-in-variables in binary regression—Berkson case. *Journal of the American Statistical Association* **83** 739–743.

CARROLL, R. J. (1989). Covariance analysis in generalized linear measurement error models. *Statistics in Medicine* **8** 1075–1093.

CARROLL, R. J., GALLO, P. P. and GLESER, L. J. (1985). Comparison of least squares and errors-in-variables regression, with special reference to randomized analysis of covariance. *Journal of the American Statistical Association* **80** 929-932.

CARROLL, R. J. and STEFANSKI, L. A. (1990). Approximate quasilikelihood estimation in models with surrogate predictors. *Journal of the American Statistical Association* **85** 652-663.

CARROLL, R. J., SPIEGELMAN, C. H., LAN, K. K. G., BAILEY, K. T. and ABBOTT, R. D. (1984). On errors-in-variables for binary regression models. *Biometrika* **71** 19–25.

CARROLL, R. J., KÜCHENHOFF, H., LOMBARD, F. and STEFANSKI, L. A. (1996). Asymptotics for the SIMEX estimator in nonlinear measurement error models. *Journal of the American Statistical Association* **91** 242–250.

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. and CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models: A modern Perspective*, 2 ed. Chapman & Hall, Boca Raton.

CLAYTON, D. (1992). Models for the Longitudinal Analysis of Cohort and Case-Control Studies with Inaccurately Measured Exposures. In *Statistical Models for Longitudinal Studies of Health* (J. H. Dwyer, M. Feinleib, P. Lippert and H. Hoffmeister, eds.) 301–331. Oxford University Press.

COOK, J. R. and STEFANSKI, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* **89** 1314-1328.

CORNFIELD, J. (1962). Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: A discriminant function analysis. *Federal Proceedings* **21** 58-61.

DELLAPORTAS, P. and STEPHENS, D. A. (1995). Analysis of errors-in-variables regression models. *Biometrics* **51** 1085-1095.

DEVANARAYAN, V. and STEFANSKI, L. A. (2002). Empirical simulation extrapolation for measurement error models with replicate measurements. *Statistics & Probability Letters* **59** 219–225.

FONG, Y., RUE, H. and WAKEFIELD, J. (2010). Bayesian inference for generalized linear mixed models. *Biostatistics* **11** 397–412.

FULLER, W. A. (1987). *Measurement Error Models*. John Wiley & Sons, New York.

GAMERMAN, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing* **7** 57–68.

GELFAND, A. E., SAHU, S. K. and CARLIN, B. P. (1995). Efficient parametrisations for normal linear mixed models. *Biometrika* **82** 479–488.

GELFAND, A., SAHU, S. K. and CARLIN, B. P. (1996). Efficient parametrizations for generalized linear mixed models. (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 48–74.

GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85** 398–409.

GLESER, L. J. (1990). Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models. In *Statistical Analysis of Measurement Error Models and Applications* (P. J. Brown and W. A. Fuller, eds.) American Mathematical Society.

GLESER, L. J., CARROLL, R. J. and GALLO, P. P. (1987). The limiting distribution of least squares in an errors-in-variables regression model. *Annals of Statistics* **15** 220-233.

GUSTAFSON, P. (2004). *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*. Chapman & Hall/CRC, Boca Raton.

KANNEL, W. B., NEATON, J. B. and WENTWORTH, D. (1986). Overall and coronary heart-disease mortality-rates in relation to major risk-factors in 325'348 men screened for the MRFIT. *American Heart Journal* **112:4** 825-836.

KERBER, R. A., TILL, J. E., SIMON, S. L., LYON, J. L., THOMAS, D. C., PRESTON-MARTIN, S., RALLISON, M. L., LLOYD, R. D. and STEVENS, W. (1993). A cohort study of thyroid disease in relation to fallout from nuclear weapons testing. *Journal of the American Medical Association* **270** 2076–2082.

KÜCHENHOFF, H., MWALILI, S. M. and LESAFFRE, E. (2006). A general method for dealing with misclassification in regression: the misclassification SIMEX. *Biometrics* **62** 85–96.

Lederer, W. and Küchenhoff, H. (2009). simex: SIMEX- and MCSIMEX-Algorithm for measurement error models http://CRAN.R-project.org/package=simex.

Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing* **10** 325–337.

Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009). The BUGS project: Evolution, critique, and future directions. *Statistics in Medicine* **28** 3049–3067.

MacMahon, S., Peto, R., Cutler, J., Collins, R., Sorlie, P., Neaton, J., Abbott, R., Godwin, J., Dyer, A. and Stamler, J. (1990). Blood pressure, stroke and coronary heart disease: Part 1, prolonged differences in blood pressure: Prospective observational studies corrected for the regression dilution bias. *Lancet* **335** 765-774.

Martin, A. D., Quinn, K. M. and Park, J. H. (2011). MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software* **42** 22.

Martino, S. and Rue, H. (2010). Case Studies in Bayesian Computation using INLA. In *Complex Data Modeling and Computationally Intensive Statistical Methods* (P. Mantovan and P. Secchi, eds.) 99–114. Springer.

Martins, T. G. and Rue, H. (2012). Extending INLA to a class of near-Gaussian latent models. *ArXiv e-prints*. arXiv:1210.1434.

Martins, T. G., Simpson, D., Lindgren, F. and Rue, H. (2012). Bayesian computing with INLA: new features. *ArXiv e-prints*. arXiv:1210.0333.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, second ed. *Monographs on Statistics and Applied Probability* 37. Chapman and Hall, New York.

Paine, T. C. E., Stenflo, M., Philipson, C. D., Saner, P., Bagchi, R., Ong, R. C. and Hector, A. (2012). Differential growth responses in seedlings of ten species of Dipterocarpaceae to experimental shading and defoliation. *Journal of Tropical Ecology* **28** 377–384.

Paul, M., Riebler, A., Bachmann, L. M., Rue, H. and Held, L. (2010). Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations. *Statistics in Medicine* **29** 1325–1339.

Pearson, K. (1902). On the mathematical theory of errors of judgement. *Philosophical Transactions of the Royal Society of London A* **198** 235-299.

R Core Team, (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Richardson, S. and Gilks, W. R. (1993). Conditional independence models for epidemiological studies with covariate measurement error. *Statistics in Medicine* **12** 1703-1722.

Riebler, A., Held, L. and Rue, H. (2012). Estimation and extrapolation of time trends in registry data - Borrowing strength from related populations. *Annals of Applied Statisticts* **6** 304-333.

Roos, M. and Held, L. (2011). Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *Bayesian Analysis* **6** 259–278.

Rudemo, M., Ruppert, D. and Streibig, J. C. (1989). Random-effect models in nonlinear regression with applications to bioassay. *Biometrics* **45** 349–362.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society - Series B* **71** 319-392.

Ruiz-Cárdenas, R., Krainski, E. T. and Rue, H. (2012). Direct fitting of dynamic models using integrated nested Laplace approximations–INLA. *Computational Statistics & Data Analysis* **56** 1808—1828.

Schrödle, B., Held, L. and Rue, H. (2012). Assessing the impact of a movement network on the spatiotemporal spread of infectious diseases. *Biometrics* **68** 736—744.

Schrödle, B., Held, L., Riebler, A. and Danuser, J. (2011). Using INLA for the evaluation of veterinary surveillance data from Switzerland: A case study. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **60** 261-279.

Simon, S. L., Till, J. E., Lloyd, R. D., Kerber, R. L., Thomas, D. C., Preston-Martin, S., Lyon, J. L. and Stevens, W. (1995). The Utah leukemia case-control study: dosimetry methodology and results. *Health Physics* **68** 460–471.

Statisticat, and LLC. , (2013). LaplacesDemon: Complete Environment for Bayesian Inference R package version 13.02.04.

Stephens, D. A. and Dellaportas, P. (1992). Bayesian analysis of generalised linear models with covariate measurement error. In *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger,

A. P. Dawid and A. F. M. Smith, eds.) Oxford Univ Press.

Subar, A. F., Thompson, F. E., Kipnis, V., Midthune, D., Hurwitz, P., McNutt, S., McIntosh, A. and Rosenfeld, S. (2001). Comparative Validation of the Block, Willett, and National Cancer Institute Food Frequency Questionnaires: The Eating at America's Table Study. *American Journal of Epidemiology* **154** 1089-1099.

Wald, A. (1940). The fitting of straigth lines if both variables are subject to error. *Annals of Mathematical Statistics* **11** 284-300.

West, M. (1985). Generalized linear models: scale parameters, outlier accommodation and prior distributions. In *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 531–558. North-Holland, Amsterdam.

SUPPLEMENTARY MATERIAL

7. R-CODE FOR THE THREE APPLICATIONS IN THE MAIN TEXT

In this section we guide the reader through the `r-inla` code and technical details of the three examples discussed in the main text. On www.r-inla.org/examples/case-studies/muff-etal-2013 selected data and R-code are provided for download. The `r-inla` package can be installed by typing the following command line in the `R` terminal:

```
source("http://www.math.ntnu.no/inla/givemeINLA.R")
upgrade.inla(testing=TRUE)
```

Using

```
inla.version()
```

information regarding the actual installed version is shown. Here, we used the `r-inla` version built on Feb 10 2013. For more information regarding the installation process we refer to www.r-inla.org.

7.1 Inbreeding in Swiss ibex populations

The object `data` consists of three columns:

$$\boldsymbol{y} \quad \boldsymbol{w} \quad \textbf{error.prec}$$

They contain (for $n = 26$):

- $y_1 \dots y_n$: The populations' intrinsic growth rates.
- $w_1 \dots w_n$: The estimated inbreeding coefficients.
- $\text{error.prec}_1 \dots \text{error.prec}_n$: The error precisions in the estimates $\boldsymbol{w}$.

Start with the prior specification process, as described in the main text:

```
attach(data)
varYEst <- (summary(lm(y~w))$sigma^2)/2
varXEst <- var(w) - mean(1/error.prec)
prec.y <- 1/varYEst
prec.x <- 1/varXEst
prec.u <- 1

prior.prec.y <- c(10, 9/prec.y) # prior mode at prec.y
prior.prec.x <- c(10, 9/prec.x) # prior mode at prec.x
prior.prec.u <- c(2, 1)
prior.beta <- c(0, 0.0001)
```

Next, we define the INLA model formula, where the new `mec` model is employed. Note that the heteroscedasticity in the error is encoded by assigning the vector of error precisions `error.prec` to the `scale` option. The model contains four hyperparameters:

- `beta` corresponds to β_x, the slope coefficient of the error-prone covariate $\boldsymbol{x}$, with a Gaussian prior.
- `prec.u` is the error precision τ_u with gamma prior.
- `prec.x` is the precision τ_x of $\boldsymbol{x} \sim \mathcal{N}(\alpha_0 \mathbf{1}, \tau_x \mathbf{I})$ with gamma prior.

- `mean.x` corresponds to the mean α_0, which is fixed here at 0 due to covariate centering.

The prior settings are defined in the different entries of the list `hyper`. The option `fixed` specifies whether the corresponding quantity should be estimated or fixed at the `initial` value. The field `param` captures the prior parameters of the corresponding prior distribution. Gaussian prior distributions are the default for `beta` and `mean.x`, while log-gamma distributions are used for the log-transformed precisions `prec.u` and `prec.x`. Note hereby, that if a variable τ is gamma distributed with shape parameter a and rate parameter b leading to the mean a/b and variance a/b^2, then $\log(\tau)$ is log-gamma distributed with the same parameters a and b.

```
library(INLA)
formula <- y ~ f(w, model = "mec", scale = error.prec, hyper = list(
   beta = list(
     param = prior.beta,
     fixed = FALSE
     ),
   prec.u = list(
     param = prior.prec.u,
     initial = log(prec.u),
     fixed = FALSE
     ),
   prec.x = list(
     param = prior.prec.x,
     initial = log(prec.x),
     fixed = FALSE
     ),
   mean.x = list(
     initial = 0,
     fixed = TRUE
     )
   )
 )
```

The call of the `inla` function includes the specifications for τ_ε, the hyperparameter of the Gaussian regression model. These can be controlled via the `control.family` option. The prior distribution for the intercept β_0 is specified in the `control.fixed` option.

```
r <- inla(formula, data = data.frame(y, w, error.prec),
          family = "gaussian",
          control.family = list(
              hyper = list(
                 prec = list(param = prior.prec.y,
                          initial = log(prec.y),
                          fixed = FALSE
                 )
              )
           ),
```

```
            control.fixed = list(
              mean.intercept = prior.beta[1],
              prec.intercept = prior.beta[2]
            )
      )
r <- inla.hyperpar(r, dz = 0.5, diff.logdens = 20)
```

The last command improves the estimates of the posterior marginals for the hyperparameters of the model. The call is optional, but a slightly better agreement with the MCMC posterior marginals was found in this example. To get a quick overview of the results, use the `summary` command.

```
summary(r)
```

7.2 Influence of systolic blood pressure on coronary heart disease

The object `data` consists of four columns:

$$\boldsymbol{y} \quad \boldsymbol{w_1} \quad \boldsymbol{w_2} \quad \boldsymbol{z}$$

They contain (for $n = 641$):

- $y_1 \dots y_n$: The binary response $y_i \in \{0, 1\}$.
- $w_{11} \dots w_{1n}$: $\log(\text{SBP} - 50)$ at examination 1.
- $w_{21} \dots w_{2n}$: $\log(\text{SBP} - 50)$ at examination 2.
- $z_1 \dots z_n$: Smoking status $z_i \in \{0, 1\}$.

As described in the main text, the hierarchical model of this example is formulated in INLA as a joint model by applying the `copy` feature. The full model can be written as

$$\underbrace{\begin{bmatrix} y_1 & \texttt{NA} & \texttt{NA} \\ \vdots & \vdots & \vdots \\ y_n & \texttt{NA} & \texttt{NA} \\ \texttt{NA} & 0 & \texttt{NA} \\ \vdots & \vdots & \vdots \\ \texttt{NA} & 0 & \texttt{NA} \\ \texttt{NA} & \texttt{NA} & \overline{w}_{1\cdot} \\ \vdots & \vdots & \vdots \\ \texttt{NA} & \texttt{NA} & \overline{w}_{n\cdot} \end{bmatrix}}_{\texttt{Y}} = \beta_0 \underbrace{\begin{bmatrix} 1 \\ \vdots \\ 1 \\ \texttt{NA} \\ \vdots \\ \texttt{NA} \\ \texttt{NA} \\ \vdots \\ \texttt{NA} \end{bmatrix}}_{\texttt{beta.0}} + \beta_x \underbrace{\begin{bmatrix} 1 \\ \vdots \\ n \\ \texttt{NA} \\ \vdots \\ \texttt{NA} \\ \texttt{NA} \\ \vdots \\ \texttt{NA} \end{bmatrix}}_{\texttt{beta.x}} + \underbrace{\begin{bmatrix} \texttt{NA} \\ \vdots \\ \texttt{NA} \\ -1 \\ \vdots \\ -n \\ 1 \\ \vdots \\ n \end{bmatrix}}_{\texttt{idx.x}} + \beta_z \underbrace{\begin{bmatrix} z_1 \\ \vdots \\ z_n \\ \texttt{NA} \\ \vdots \\ \texttt{NA} \\ \texttt{NA} \\ \vdots \\ \texttt{NA} \end{bmatrix}}_{\texttt{beta.z}} + \alpha_0 \underbrace{\begin{bmatrix} \texttt{NA} \\ \vdots \\ \texttt{NA} \\ 1 \\ \vdots \\ 1 \\ \texttt{NA} \\ \vdots \\ \texttt{NA} \end{bmatrix}}_{\texttt{alpha.0}} + \alpha_z \underbrace{\begin{bmatrix} \texttt{NA} \\ \vdots \\ \texttt{NA} \\ z_1 \\ \vdots \\ z_n \\ \texttt{NA} \\ \vdots \\ \texttt{NA} \end{bmatrix}}_{\texttt{beta.z}} .$$

The reader is guided through the `r-inla` code for this joint model formulation in the following. The terms below the brackets indicate the names as they will be employed in the code. Start with the prior specification process, as described in the main text:

```
attach(data)
w <- rowMeans(cbind(w1, w2)) #the mean of w1 and w2
```

```
n <- nrow(data)  #641
Ntrials <- rep(1, n)

varUEst <- sum((w1 - w)^2 + (w2 - w)^2)/n
varXEst <- summary(lm(w~z))$sigma^2 - varUEst
prec.u <- 1/varUEst
prec.x <- 1/varXEst

prior.prec.u <- c(n, 1/2*sum((w1 - w)^2 + (w2 - w)^2))
prior.prec.x <- c(100, 99/prec.x) # prior mode at prec.x
prior.beta <- c(0, 0.0001)
```

Second, the response matrix `Y` and the data vectors are filled according to the naming of the above joint model equation:

```
Y <- matrix(NA, 3*n, 3)
Y[1:n, 1] <- y
Y[n+(1:n), 2] <- rep(0, n)
Y[2*n+(1:n), 3] <- w
beta.0 <- c(rep(1, n), rep(NA, n), rep(NA, n))
beta.x <- c(1:n, rep(NA, n), rep(NA, n))
idx.x <- c(rep(NA, n), 1:n, 1:n)
weight.x <- c(rep(1, n), rep(-1, n), rep(1, n))
beta.z <- c(z, rep(NA, n), rep(NA, n))
alpha.0 <- c(rep(NA, n), rep(1, n), rep(NA, n))
alpha.z <- c(rep(NA, n), z, rep(NA, n))
data.joint <- data.frame(Y, beta.0, beta.x, idx.x, weight.x,
     beta.z, alpha.0, alpha.z)
```

The next step contains the definition of the INLA formula. There are four fixed effects (β_0, β_z, α_0 and α_z) and two random effects. The latter are needed to encode for the fact that the values of $\boldsymbol{x}$ in the exposure (7) and error model (8) are assigned the same values as in the regression model (6), where $\beta_x\boldsymbol{x}$ represents a product of two unknown quantities. The two random effects terms are:

- `f(beta.x,...)`: The `copy="idx.x"` call guarantees the assignement of identical values to $\boldsymbol{x}$ in all components of the joint model. As discussed in the main text, β_x is treated as a hyperparameter, namely the scaling parameter of the copied process $\boldsymbol{x}^\star$. Note that the initial value of 1.91 given in the following corresponds to the naive estimate of β_x.
- `f(idx.x,...)` : `idx.x` contains the $\boldsymbol{x}$ values, encoded as an i.i.d. Gaussian random effect, and weighted with `weight.x` to ensure correct signs in the joint model. The `values` option contains the vector of all values assumed by the covariate for which the effect is estimated. It must be a numeric vector, a vector of factors or NULL. The precision `prec` of the random effect is fixed at $\tau = \exp(-15)$.

```
library(INLA)
formula <- Y ~  f(beta.x, copy = "idx.x",
     hyper = list(beta = list(initial = 1.91,
       param = prior.beta, fixed = FALSE))) +
   f(idx.x, weight.x, model = "iid", values = 1:n,
```

```
    hyper = list(prec = list(initial = -15, fixed = TRUE))) +
  beta.0 - 1 + beta.z + alpha.0 + alpha.z
```

Since there is no common intercept in the joint model, it has to be explicitly removed using `-1`. The call of the `inla` function is given next. The following options need some explanation:

- `family`: There are three different likelihoods here, namely the binomial likelihood of the regression model and two Gaussian likelihoods, one for the exposure and one for the error model. They correspond to the different columns in the response matrix `Y`.
- `control.family`: Specification of the hyperparameters for the three likelihoods, in the same order as given in `family`. The binomial likelihood does not contain any hyperparameters, thus the respective list is empty. In the second and third likelihood the hyperparameters τ_x and τ_u need to be specified.
- `control.fixed`: Prior specification for the fixed effects.
- `control.inla`: Specification of internal parameters used in the optimization routine of INLA. This is optional and not necessary if covariates were centered. For details of the options consult `?control.inla`.

```
r <- inla(formula, Ntrials = Ntrials, data = data.joint,
          family = c("binomial", "gaussian", "gaussian"),
          control.family = list(
                  list(hyper = list()),
                  list(hyper = list(
                              prec = list(initial = log(prec.x),
                              param = prior.prec.x,
                              fixed = FALSE))),
                  list(hyper = list(
                              prec = list(initial=log(prec.u),
                              param = prior.prec.u,
                              fixed = FALSE)))),
          control.fixed = list(
                  mean.intercept = prior.beta[1],
                  prec.intercept = prior.beta[2],
                  mean = prior.beta[1],
                  prec = prior.beta[2]),
          control.inla = list(
                  h = 1e-5,
                  tolerance = 1e-6,
                  int.strategy = "grid",
                  dz = 0.2)
       )
summary(r)
```

7.3 Seedling growth accross different light conditions

Analysis with the `meb` model

The object `data` consists of the three columns:

$$\boldsymbol{y} \quad \boldsymbol{w} \quad \boldsymbol{z}$$

They contain (for $n = 60$):

- $y_1 \dots y_n$: The number of new leaves.
- $w_1 \dots w_n$: log(%light) for the target light intensities under dark, middle and light conditions (i.e., only three different values). Here, the centered values are used.
- $z_1 \dots z_n$: Degree of defoliation (0%, 25%, 50%, 75%, respectively the centered values).

Let us start again with prior specification process in accordance to the main text:

```
attach(data)
n <- 60           # number of seedlings
s <- 15           # number of shadehouses
w <- w + rep(rnorm(s,0,1e-4),each=n/s)
individual <- 1:n # id to incorporate individual random effects

prec.u <- 4.98        # estimated value
prec.tau <- 57.83     # estimated value
prior.beta <- c(0,0.0001)
prior.prec.u <- c(100,99/prec.u)  # prior mode at prec.u
prior.tau <- c(100,99/prec.tau)   # prior mode at prec.tau
```

The fourth line contains a trick to ensure that the light values `w` from the $s = 15$ shadehouses are not completely identical, because in the new `meb` model only the unique values of `w` are used. Thus, if two or more elements of `w` are *identical*, then they refer to the *same* element in the covariate $\boldsymbol{x}$. Next, we define the `meb` model formula. The model contains two hyperparameters:

- `beta` corresponds to β_x, the slope coefficient of the error-prone covariate $\boldsymbol{x}$, with a Gaussian prior.
- `prec.u` is the error precision τ_u with gamma prior.

The prior settings are defined in the different entries of the list `hyper`. The option `fixed` specifies whether the corresponding quantity should be estimated or fixed at the `initial` value. The field `param` captures the prior parameters of the corresponding prior distribution. A Gaussian prior distribution is the default for `beta`, while a gamma distribution is used for `prec.u` (again defined as log-gamma distribution for the log-precision).
The model contains as additional fixed effect the degree of defoliation `z`, plus an additional i.i.d. random effects term per individual to account for unspecified heterogeneity, specified in `f(individual,...)`, which extends the GLM to a GLMM:

```
library(INLA)
formula <- y ~  f(w, model="meb", hyper = list(
             beta = list(
                       param = prior.beta,
                       fixed = FALSE
             ),
             prec.u = list(
                       param = prior.prec.u,
                       initial = log(prec.u),
                       fixed = FALSE
             )
   )) +
   z +
   f(individual, model = "iid", values = 1:n, hyper = list(prec = list(
             initial = log(prec.tau),
             param = prior.tau
         )
      )
   )
```

The call of the `inla` function includes the specification of the family, which is Poisson here and thus includes no additional hyperparameters. The prior distributions for the intercept β_0 and the slope β_z are specified in the `control.fixed` option.

```
r <- inla(formula, data = data.frame(y, w, z, individual),
          family = "poisson",
          control.fixed = list(
                    mean.intercept = prior.beta[1],
                    prec.intercept = prior.beta[2],
                    mean = prior.beta[1],
                    prec = prior.beta[2]),
     )
summary(r)
```

Analysis with the `copy` feature

As described in the main text, as an alternative to the use of the new `meb` model, the same results can be obtained by employing the `copy` feature in INLA. The approach is similar to the one taken in Section 7.2.

The object `data` now contains an additional fourth column:

$\boldsymbol{y}$	$\boldsymbol{w}$	$\boldsymbol{z}$	**sh**

Column **sh** contains the values $\text{sh}_1, \ldots, \text{sh}_n$, where sh_i is the index of the shadehouse where seedling i was grown, $i = 1, \ldots, 15$. As the error model in this example is Berkson, the joint model simplifies to two equations and the response matrix has only two columns. The model can be represented as

$$
\underbrace{\begin{bmatrix} y_1 & \texttt{NA} \\ \vdots & \vdots \\ y_n & \texttt{NA} \\ \texttt{NA} & -w_1 \\ \vdots & \vdots \\ \texttt{NA} & -w_s \end{bmatrix}}_{\texttt{Y}} = \beta_0 \underbrace{\begin{bmatrix} 1 \\ \vdots \\ 1 \\ \texttt{NA} \\ \vdots \\ \texttt{NA} \end{bmatrix}}_{\texttt{beta.0}} + \beta_x \underbrace{\begin{bmatrix} \mathrm{sh}_1 \\ \vdots \\ \mathrm{sh}_n \\ \texttt{NA} \\ \vdots \\ \texttt{NA} \end{bmatrix}}_{\texttt{beta.x}} + \underbrace{\begin{bmatrix} \texttt{NA} \\ \vdots \\ \texttt{NA} \\ -1 \\ \vdots \\ -s \end{bmatrix}}_{\texttt{idx.x}} + \beta_z \underbrace{\begin{bmatrix} z_1 \\ \vdots \\ z_n \\ \texttt{NA} \\ \vdots \\ \texttt{NA} \end{bmatrix}}_{\texttt{beta.z}} + \underbrace{\begin{bmatrix} 1 \\ \vdots \\ n \\ \texttt{NA} \\ \vdots \\ \texttt{NA} \end{bmatrix}}_{\texttt{gamma}} . \tag{15}
$$

Note that there are $n = 60$ seedlings distributed over $s = 15$ shadehouses, whereas always five shadehouses belong to the same light condition (dark, middle, light). There are thus 15 different correct light intensities ($\boldsymbol{x}$, one value per shadehouse), but only 3 different target light intensities ($\boldsymbol{w}$, one value per light condition). Terms below the brackets correspond to the names in the R-code.
Let us start again with prior specification process in accordance to the main text:

```
attach(data)
w.red <- aggregate(w, by = list(sh), FUN = mean)[,2]
n <- 60          # number of seedlings
s <- 15          # number of shadehouses

prec.u <- 4.98         # estimated value
prec.tau <- 57.83      # estimated value
prior.beta <- c(0,0.0001)
prior.prec.u <- c(100,99/prec.u)  # prior mode at prec.u
prior.tau <- c(100,99/prec.tau)   # prior mode at prec.tau
```

The `aggregate` command in the second line aggregates the vector $\boldsymbol{w}$ of length $n = 60$ into the 15 (one per shadehouse) unique light values.
Next, the response matrix `Y` and the data vectors are filled according to the naming of Equation (15):

```
Y <- matrix(NA, n+s, 2)
Y[1:n, 1] <- y
Y[n+(1:s), 2] <- -w.red
beta.0 <- c(rep(1, n), rep(NA, s))
beta.x <- c(sh, rep(NA, s))
idx.x <- c(rep(NA, n), 1:s)
weight.x <- c(rep(NA, n), -rep(1, s))
beta.z <- c(z, rep(NA, s))
gamma <- c(1:n, rep(NA, s))
data.joint <- data.frame(Y, beta.0, beta.x, idx.x, weight.x, beta.z, gamma)
```

The definition of the INLA formula is almost analogous to the one in Section 7.2. The main difference is the additional i.i.d. random effects term per individual γ_{ijk}, specified in `f(gamma,...)`, which extends the GLM to a GLMM:

```
library(INLA)
formula <- Y ~  beta.0 - 1 +
```

```
    f(beta.x, copy = "idx.x",
          hyper = list(beta = list(param = prior.beta, fixed = FALSE))) +
    f(idx.x, weight.x, model = "iid", values = 1:s,
          hyper = list(prec = list(initial = -15, fixed = TRUE))) +
    beta.z +
    f(gamma, model = "iid", values = 1:n,
          hyper = list(prec = list(initial = log(prec.tau), param = prior.tau)))
```

As in Section 7.2 we have to explicitly remove the common intercept using `-1`. The call of the INLA function is as well in analogy to Section 7.2, but there are only two likelihoods involved here: the Poisson likelihood for the regression model and the Gaussian likelihood for the error model. The former has no additional hyperparameters, while in the latter the error precision τ_u needs specification.

```
r <- inla(formula, data = data.joint,
          family = c("poisson", "gaussian"),
          control.family = list(
                    list(hyper = list()),
                    list(hyper = list(
                              prec = list(initial=log(prec.u),
                              param = prior.prec.u,
                              fixed = FALSE)))),
          control.fixed = list(
                    mean.intercept = prior.beta[1],
                    prec.intercept = prior.beta[2],
                    mean = prior.beta[1],
                    prec = prior.beta[2])
       )
summary(r)
```

8. SUPPLEMENTS TO SECTION 5.2 IN THE MAIN TEXT

8.1 Full conditionals for the MCMC sampler

Let all variables be defined as in Section 5.2 of the main text. The full conditionals for the unknowns in the regression model are given as follows:
For $\boldsymbol{\beta} = (\beta_0, \beta_x, \beta_z)^\top$ we have

$$\begin{aligned} \boldsymbol{\beta} \,|\, rest &\propto \pi(\boldsymbol{y} \,|\, \boldsymbol{x}, \boldsymbol{z}) \cdot \pi(\boldsymbol{\beta}) \\ &\propto \exp\left(\sum_{i=1}^{n} y_i \eta_i - \sum_{i=1}^{n} \log(1 + e^{\eta_i}) - \frac{\tau_{\alpha\beta}}{2} \boldsymbol{\beta}^\top \boldsymbol{\beta} \right) , \end{aligned}$$

where $\eta_i = \beta_0 + \beta_x x_i + \beta_z z_i$ and $\tau_{\alpha\beta} = 10^{-4}$. The $\boldsymbol{\alpha} = (\alpha_0, \alpha_z)^\top$ coefficients can be sampled from a Gaussian distribution. Let D bet the matrix with rows $D_i^\top := (1 \;\; z_i^\top)$ and $\Delta := \tau_{\alpha\beta}/\tau_x$. Then

$$\begin{aligned} \boldsymbol{\alpha} \,|\, rest &\propto \pi(\boldsymbol{x} \,|\, rest) \cdot \pi(\alpha) \\ &\propto \exp\left(-\frac{\tau_x}{2} (\boldsymbol{x} - D\boldsymbol{\alpha})^\top (\boldsymbol{x} - D\boldsymbol{\alpha}) - \frac{\tau_{\alpha\beta}}{2} \boldsymbol{\alpha}^\top \boldsymbol{\alpha} \right) \\ &\sim \mathcal{N}\left((D^\top D + \Delta \mathbf{I})^{-1} D^\top \boldsymbol{x}, \tau_x (D^\top D + \Delta \mathbf{I}) \right) , \end{aligned}$$

where the second argument in the last expression is again the precision. To sample from the distribution of the latent variable $\boldsymbol{x}$, full conditionals for x_i are needed:

$$\begin{aligned} x_i \,|\, rest &\propto \pi(y_i \,|\, x_i, z_i) \cdot \pi(w_i \,|\, x_i) \cdot \pi(x_i \,|\, z_i) \\ &\propto \exp\left(y_i \eta_i - \log(1 + e^{\eta_i}) - \frac{\tau_u}{2} (w_i - x_i)^2 - \frac{\tau_x}{2} (x_i - \alpha_0 - \alpha_z z_i)^2 \right) . \end{aligned}$$

Finally, the precisions can be sampled from gamma distributions

$$\begin{aligned} \tau_x \,|\, rest &\propto \pi(\boldsymbol{x} \,|\, \boldsymbol{z}) \cdot \pi(\tau_x) \\ &\sim \mathrm{G}\left(a_x + \frac{n}{2}, b_x + \frac{1}{2} (\boldsymbol{x} - D\boldsymbol{\alpha})^\top (\boldsymbol{x} - D\boldsymbol{\alpha}) \right) , \end{aligned}$$

and

$$\begin{aligned} \tau_u \,|\, rest &\propto \pi(\boldsymbol{w} \,|\, \boldsymbol{x}) \cdot \pi(\tau_u) \\ &\sim \mathrm{G}\left(a_u + \frac{n}{2}, b_u + \frac{1}{2} (\boldsymbol{w} - \boldsymbol{x})^\top (\boldsymbol{w} - \boldsymbol{x}) \right) . \end{aligned}$$

8.2 MCMC and INLA posterior marginals

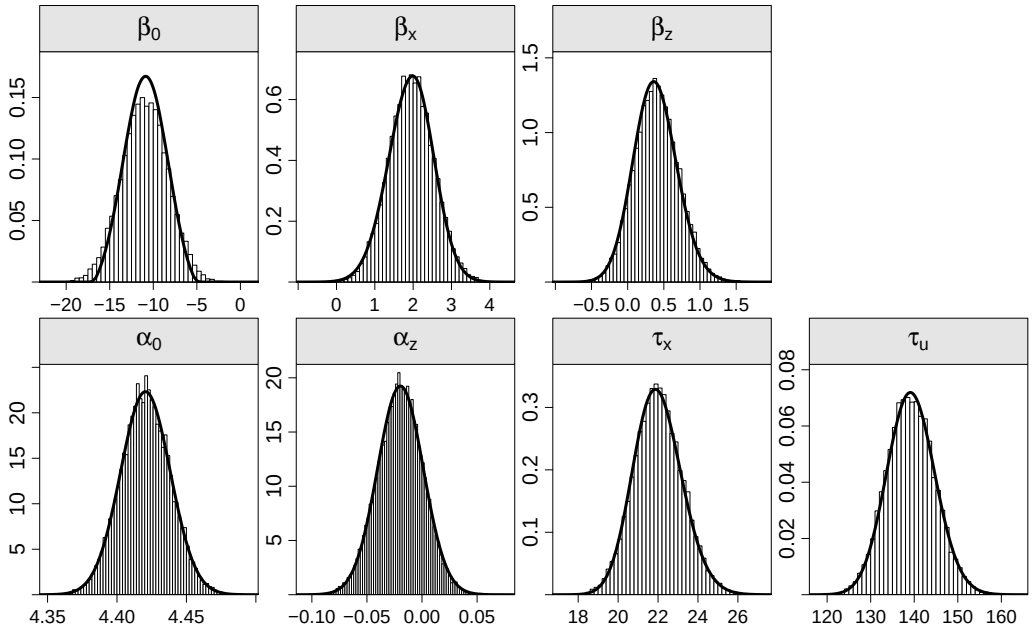

FIG 9. *Comparison of the MCMC samples (histograms) with the INLA posterior marginals (lines) for the Framingham data.*